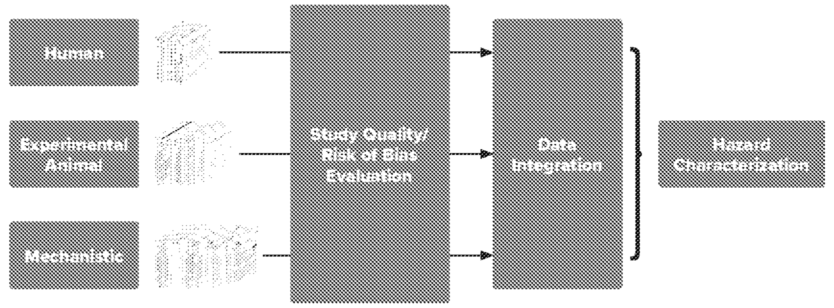
ToxStrategies

Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws L[1]

[1] ToxStrategies, Austin, TX 78759. [2] ToxStrategies, Asheville, NC 28804. [3] ToxStrategies, Mission Viejo, CA 92691. [4] ToxStrategies, Tallahassee, FL 32309. [5] ToxStrategies, Katy, TX 77494

Poster #: P305
Abstract #: 2801



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2018)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characerization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:*** 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

* OHAT RoB guidelines encourages reviewers to expand/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments
Epidemiology	10	9
Animal Toxicology	11	12
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study] In vivo: 5 In vitro (cell culture): 26 In vitro (in ovo): 21 In vitro (ex ovo): 3 In vitro (ex vivo): 7 In vitro (zebrafish): 5 Unknown model: 1
Total	43	89

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier II
Lagakos et al. (1986)	Cross-sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

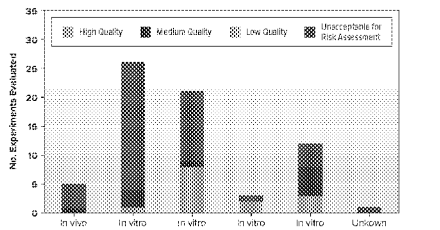


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

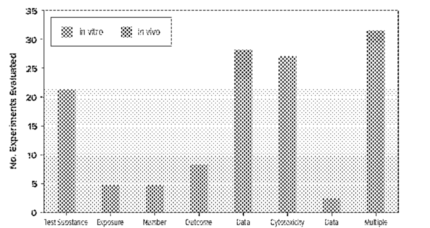
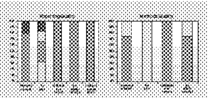
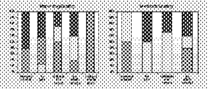
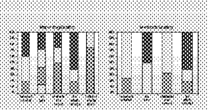


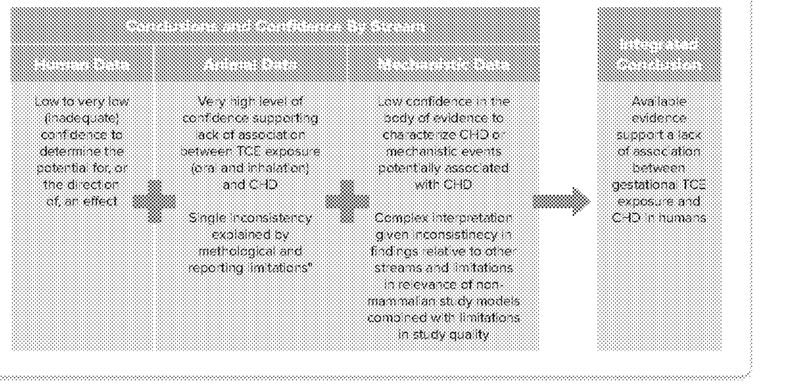
Table 6. Comparison of In Vitro Study Quality Evaluation Tools

Study	Experimental Model & Endpoint	Study Quality Evaluation Outcome: TSCA	SciRAP	
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 85 out of 100 **Consequence:** None defined	
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined	
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined	

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
    1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
    1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
    2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
    3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion



Low to very low (inadequate) confidence to determine the potential for, or the direction of, an effect

Very high level of confidence supporting lack of association between TCE exposure (oral and inhalation) and CHD

Single inconsistency explained by methological and reporting limitations*

Low confidence in the body of evidence to characterize CHD or mechanistic events potentially associated with CHD

Complex interpretation given inconsistincey in findings relative to other streams and limitations in relevance of non-mammalian study models combined with limitations in study quality

Available evidence support a lack of association between gestational TCE exposure and CHD in humans

Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References
Bove FJ. (1996). PMID: 8794538.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406847.
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288582
Narotsky MG et al. (1995). PMID: 8529816.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckart PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2018). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29357719.